FEBS 18600 FEBS Letters 409 (1997) 237–241

# Identification of a novel nuclear localization signal in Sam68

Takao Ishidate, Satoshi Yoshihara, Yoshihiro Kawasaki, Badal Chandra Roy, Kumao Toyoshima[1], Tetsu Akiyama*

*Department of Oncogene Research, Institute for Microbial Diseases, Osaka University, 3-1, Yamadaoka, Suita, Osaka 565, Japan*

Received 13 March 1997; revised version received 11 April 1997

**Abstract Sam68, a nuclear RNA binding protein, binds to Src and is phosphorylated at tyrosine residues in an M-phase specific manner. Here we identified a stretch of 24 amino acid residues in the COOH-terminal portion of Sam68 which function as a nuclear localization signal. This signal sequence bears no apparent homology to any other known nuclear localization sequence. However, this sequence was found to contain a motif, PPXXR (P, Pro; R, Arg), which is conserved in various RNA binding proteins including hnRNP proteins. Replacement of Arg in this motif with Ala abolished the nuclear accumulation of a GFP fusion protein, suggesting that this residue is important in translocating the protein to the nucleus.**



*Key words:* Sam68; Nuclear localization signal; RNA binding protein

## 1. Introduction

A nuclear phosphoprotein Sam68 was first identified as a protein which binds to and is phosphorylated by the protein tyrosine kinase c-Src in mitotic fibroblasts [1,2]. In cells transformed with activated c-Src or v-Src, Sam68 is phosphorylated at Tyr during the M phase of the cell cycle and binds to both the SH2 and SH3 domains of Src through its phosphorylated Tyr and one of its proline-rich sequences [3]. In addition to Src, Sam68 has also been shown to interact with other signal transducing molecules, namely, Grb2 and PLCγ-1, although the precise physiological significance of these interactions is not known [4].

The tyrosine kinase activity of Src is elevated during both the G1-S and G2-M transitions of the cell cycle, and its activation has been shown to be required for fibroblasts to begin DNA synthesis upon stimulation with growth factors, and to enter mitosis [5–7]. Since Sam68 binds to and is phosphorylated by Src during mitosis, it has been considered as a candidate molecule for a transducer of the mitotic Src signal to downstream target(s).

Experimental evidence has revealed that a cDNA previously reported to encode the GAP-associated protein p62 actually encodes Sam68 [8]. The predicted protein product of the Sam68 cDNA contains, at the center of the molecule, a KH domain [9], recently described as an RNA binding motif which is conserved in a number of RNA binding proteins such as hnRNP-K [10], GRP33 [11], fragile X mental retardation gene product FMR-1 [12] and the *Caenohabditis elegans* germ-line-specific tumor suppressor GLD-1 [13]. The $NH_2$-terminal part of this protein also contains a region of similarity to the RGG box [14], another hallmark of RNA binding proteins. These structural characteristics suggest that Sam68 is an RNA binding protein and may play a role in mRNA metabolism, maturation and/or transport. Indeed, it has recently been shown that Sam68 can bind to poly(U) RNA homopolymers in vitro and that this binding is regulated by tyrosine phosphorylation of the protein [15]. Interestingly, a point mutation in one of the conserved residues in the KH domain of various RNA binding proteins has been shown to abolish the RNA binding activity. This mutation results in the development of the fragile X syndrome in the case of FMR-1 and germ line tumors in the case of GLD-1 [12,13], underscoring the functional importance of RNA binding mediated via this domain.

Apart from the RNA binding domains and the proline-rich SH3 binding domains, there are no characteristic amino acid sequences in Sam68 which would suggest a function, except for the COOH-terminal region which is relatively rich in tyrosine (16/78 ≒ 21%) and possibly contains mitotic tyrosine phosphorylation site [9]. Previous studies have shown that the tyrosine phosphorylation site which negatively regulates RNA binding activity resides within this region [15]. These findings suggest that the COOH-terminal tyrosine-rich region might be a regulatory domain of the protein. In this study, we investigated the function of the COOH-terminal region of Sam68 by constructing a mutant in which the COOH-terminal tyrosine-rich region was deleted. We unexpectedly found that this mutant Sam68 localizes to the cytoplasm when expressed in fibroblast cells. From a series of deletion analysis, we identified a stretch of 24 amino acids within the COOH-terminal region of Sam68 as a novel nuclear localization signal which has no similarity to any other known nuclear localization signals.

## 2. Materials and methods

### 2.1. Cell culture

NIH3T3 cells were cultured in DMEM supplemented with 10% calf serum.

### 2.2. Plasmid construction

The human Sam68 full-length cDNA was isolated by screening a human brain library using a partial cDNA as the hybridization probe, and the sequence confirmed after the cDNA was cloned into the pBlueScript vector. The Sam68 cDNA was then subcloned into the eukaryotic expression vectors pUHDneo and pMKIT. A COOH-terminal deletion mutant was constructed by digestion with *Nde*I followed by self-religation. To tag Sam68 at the COOH-terminus with the myc 9E10 epitope (EQKLISEEDL), a *Mlu*I site was created immediately upstream of the stop codon of the Sam68 cDNA by PCR mediated insertion, using a primer containing an *Mlu*I site. Synthetic oligonucleotides encoding the myc 9E10 epitope and a stop codon with an *Mlu*I site or a *Sal*I site at each end (5′-CGCGTGAGCA-

*Corresponding author. Fax: (81) (6) 879-8305.
E-mail: akiyama@biken.osaka-u.ac.jp

[1] *Present address:* Center for Adult Diseases, Nakamichi, Higashinari-ku, Osaka, Osaka 537, Japan.


*PII* S0014-5793(97)00455-9

GAAGCTGATTTCCGAGGAGGACCTGTAAG-3′, 5′-TCGACTTACAGGTCCTCCTCGGAAATCAGCTTCTGCTCA-3′) were phosphorylated, annealed, and then inserted into the *Mlu*I site and the unique *Sal*I site that exists in the multi-cloning site of the vector.

The pCX-GFP vector carrying the gene driven by the chicken beta-actin promoter was a kind gift from Dr. Masahito Ikawa (Osaka University). To construct various GFP-Sam68 fusion proteins, a *Mlu*I site was created immediately upstream of the stop codon of the GFP gene by PCR-mediated insertion. Then various portions of the COOH-terminal region of Sam68 were PCR amplified and cloned into the *Mlu*I site, using the primer sets as follows: C1, 5′-GCGACGCGTGCACCAGAAACATATGAA-3′, 5′-TTAACGCGTCCTGGTCCCATTCCAGTCGTC; C2, 5′-GCGACGCGTGCACCAGAAACATATGAA-3′, 5′-TTAACGCGTTCCATATGGGTGCTC-3′. For C3, C4, C5 and C5-R427A, the following oligonucleotides containing *Mlu*I sites at both ends were phosphorylated, annealed and inserted into the *Mlu*I site of the pCX-GFP vector: C3, 5′-CGCGTAAGGGAGCATACAGAGAGCACCCATATA-3′, 5′-CGCGTATATGGGTGCTCTTGTATGCTCCCTTA-3′; C4, 5′-CGCGTCCTCCTGCTAGGCCAGTGAAGGGAGCATACAGAGAGCACCCATATGGAA-3′, 5′-CGCGTTCCATATGGGGCTCTCTGTATGCTCCCTTCACTGGCCTAGCAGGAGGA-3′; C5, 5′CGCGTAGGCCGTCGCTGAAGGCCCCTCCTGCTAGGCCAGTGAAGGGAGCATACAGAGAGCACCCATATGGACGTTATA-3′, 5′-CGCGTATAACGTCCATATGGGTGCTCTCTGTATGCTCCCTTCACTGGCGCAGCAGGAGGGGCCTTCAGCGACGGCCTA-3′; C6-R427A, 5′-CGCGTGACGACTGGAATGGGACCAGGCCGTCGCTGAAGGCCCCTCCTGCTGCGCCAGTGAAGGGAGCATACAGAGAGCACCCATATGGACGTTATA-3′, 5′-CGCGTATAACGTCCATATGGGTGCTCTCTGTATGCTCCCTTCACTGGCGCAGCAGGAGGGGCCTTCAGCGACGGCCTGGTCCCATTCCAGTCGTCA-3′.

### 2.3. Immunofluorescence microscopy

NIH3T3 cells were grown on coverslips, transfected with expression plasmids using the Lipofectamine reagent (GIBCO), then cultured for 24 h. Coverslips were washed once in PBS, fixed in 3.7% formaldehyde/PBS for 1 h, treated with 100% methanol (−20°C) for 10 min, incubated with a monoclonal antibody against the myc 9E10 epitope (CALBIOCHEM) for 1 h, and then with a FITC-conjugated goat anti-mouse IgG antibody (CAPPEL) for 1 h. Coverslips were mounted on glass slides using the VECTASHEILD reagent (Vector Laboratories) and observed with a fluorescent microscope.

## 3. Results

As the first step in studying the function of Sam68, we constructed a deletion mutant of Sam68 lacking 72 amino acids (366–438) of its COOH-terminal region (del-C-ter in Fig. 1A). Since this region of the protein is very rich in tyrosine (21%), and possibly constitutes the tyrosine phosphorylation region, we speculated that it could be an important regulatory domain of the protein. The wild-type Sam68 cDNA and COOH-terminal deletion mutant cDNAs were tagged at their COOH-termini with the myc 9E10 epitope so as to allow detection of the expressed polypeptides in transfected cells by immunofluorescence microscopy. When the wild-type Sam68 cDNA was transfected into mouse NIH3T3 cells and immuno-stained with anti-myc 9E10 monoclonal antibody, predominantly nuclear staining was observed as reported previously (Fig. 1B-a) [9,22]. However, when the COOH-terminal deletion mutant was similarly transfected and

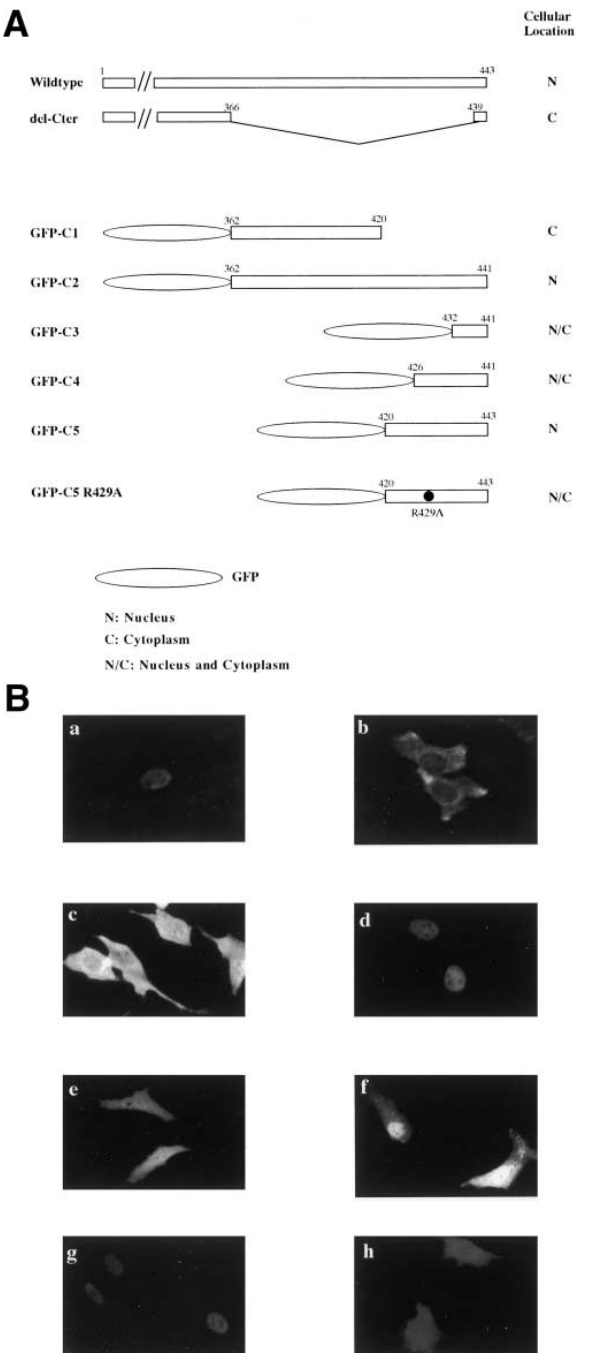
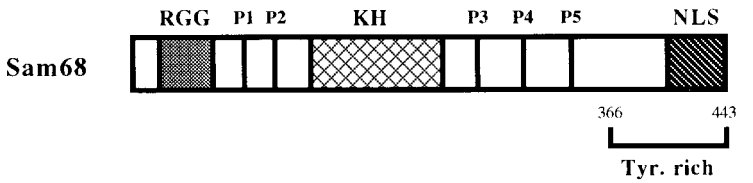



Fig. 1. Structure and subcellular localization of Sam68 and GFP-Sam68 fusion proteins. (A) Structure of Sam68 and GFP-Sam68 fusion derivatives. del-C-ter, mutant Sam68 lacking its COOH-terminal region (amino acids 366–439); GFP-C1-C5, fusion proteins consisting of GFP and various portions of the COOH-terminal region of Sam68. GFP-C5R429A, a point mutant of GFP-C5 in which Arg-429 is replaced with Ala (see Fig. 3). The oval part represents the GFP moiety while the rectangular part shows the various portions of Sam68. The numbers on top of the rectangles refer to the amino acid numbers of Sam68. The symbols N, C, N/C refer to nuclear, cytoplasmic, or nuclear and cytoplasmic localization of the fusion proteins, respectively. (B) Subcellular localization of wild-type Sam68, mutant Sam68 lacking its COOH-terminal region (amino acids 366–439) and GFP-Sam68 fusion derivatives. NIH3T3 cells were grown on coverslips and transfected with respective expression plasmids containing myc-tagged wild-type or mutant Sam68, or the various GFP-Sam68 fusion derivatives diagrammed in (A). Cells were incubated for 24 hours to allow expression of the transfected cDNAs and then fixed with 3.7% formalin and methanol. Proteins were visualized by immunofluorescence microscopy using a monoclonal antibody against the myc epitope 9E10 or a polyclonal antibody against GFP. (a) Wild type; (b) del-C-ter; (c) GFP-C1; (d) GFP-C2; (e) GFP-C3; (f) GFP-C4; (g) GFP-C5; (h) GFP.

(A)




(B)

366 YEEYGYDDTY AEQSYEGYEG YYSQSQGDSE YYDYGHGEVQ

406 DSYEAYGQDD WNGTRPSLKA PPAIPVKGAY REHPYGRY

(C)

| | | |
|---|---|---|
| Sam68 | 420 | RPSLKA PP-AR PVKGA |
| GLD-1 | 63 | WSSVRA PPPAR LLTLS |
| hnRNP-U | 194 | FAVTVA PPGAR QGQQQ |
| hnRNP-C | 141 | ARVPPP PPIAR AAVVP |
| hnRNP-K | 305 | NLPLPP PPPPR PGDLM |
| hnRNP-M4 | 282 | PKGDFF PP-ER PQQLP |
| FMR-1 | 444 | GASSRP PP-NR TDKEK |
| VP16HSV1 | 14 | ADGASP PP-PR PAGGP |
| VP16HSV2 | 12 | RDGVSP PP-PR PAGGP |

Fig. 2. Functional domains of Sam68 and the composition of the putative nuclear localization sequence. (A) The structural features of Sam68 is schematically represented. Features depicted are: RGG, two RGG boxes; KH, a KH domain; NLS, a nuclear localization sequence; P1 to P6, proline rich SH3 binding domains. (B) Amino acid sequence of the COOH-terminal region of Sam68. The region required for the nuclear localization of Sam68 is underlined, and the amino acids that are conserved in a number of RNA binding proteins are emphasized. (C) Sequence alignment of the nuclear localization signal of Sam68 with the corresponding sequences of various RNA binding proteins and viral proteins. Numbers refer to the position of the first amino acid in the primary sequence of the protein shown. Single-letter amino acid code is used. Conserved amino acids are shown at the center. Identical residues are emphasized.

stained, strong cytoplasmic staining but almost no nuclear staining was observed (Fig. 1B-b). These results suggest that the COOH-terminal 72 amino acids of Sam68 contains a signal sequence responsible for its localization in the nucleus.

To test if this stretch of 72 amino acids can serve as a nuclear localizing sequence even in a heterologous context, we fused this 72 amino-acid region to the COOH-terminus of the green fluorescent protein (GFP) derived from *Aequorea victoria* and studied the subcellular localization of the fusion proteins. GFP itself is a soluble protein with a calculated molecular weight of approximately 26.9 kDa and the amino acid sequence of GFP contains no detectable nuclear localization signal. When wild-type GFP is expressed in NIH3T3 cells and detected by immunofluorescence microscopy using the anti-GFP polyclonal antibody (Clonetech), the protein showed predominantly cytosolic localization (Fig. 1B-h). Some nuclear staining, possibly due to the passive diffusion of the protein into the nucleus, was also observed in cells where the level of GFP expression was especially high. In contrast, when a fusion protein consisting of GFP fused with the COOH-terminal 72 amino acids of Sam68 (GFP-C2 in Fig. 1A) was similarly expressed in NIH3T3, this fusion protein translocated completely to the nucleus (Fig. 1B-d). This finding further lends support to the notion that the COOH-terminal region of Sam68 contains a functional nuclear localizing sequence. Interestingly, however, there are no recognizable nuclear localizing sequences in this region, for example, the classic SV40 large T antigen type [16], the bipartite signal as found in nucleoplasmin [17], or the recently-characterized M9 sequence of hnRNP-A1 [18]. This prompted us to further delineate the sequence necessary to translocate Sam68 to the nucleus.

For this purpose, various portions of the COOH-terminal 72 amino acids of Sam68 were fused to the COOH-terminus of GFP and the subcellular localization of these fusion proteins were studied. Fusion of amino acids 362–420 (GFP-C1) to GFP had no effect on the subcellular localization of the resulting protein; it remained mainly in the cytoplasm and did not show significant accumulation in the nucleus (Fig. 1B-c). Given that amino acids 362–441 (GFP-C2) were sufficient to confer nuclear localization, we surmised that the COOH-terminal amino acids 420–441 may contain the nuclear targeting signal. Indeed, when this stretch of 24 amino acids was fused to GFP (GFP-C5), the fusion protein was efficiently targeted to the nucleus (Fig. 1B-g). Several deletions of the $NH_2$-terminus of the 24 amino acid segment (GFP-C3 and GFP-C4) abolished its nuclear targeting activity almost completely (Fig. 1B-e and -f). From these results, we conclude that the COOH-

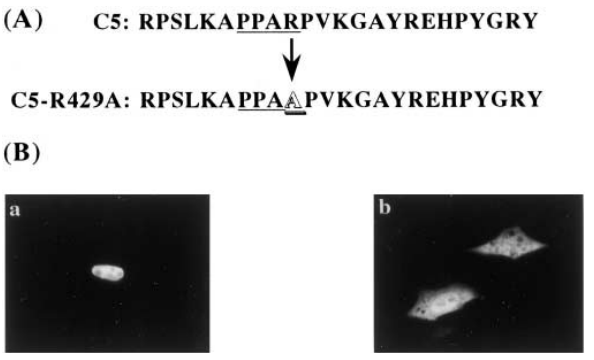



Fig. 3. Effect of point mutation in the nuclear localization sequence on the nuclear targeting activity. (A) Amino acid sequences of wild-type and mutant nuclear localization signals. C5, wild-type; C5-R429A, mutant containing Ala instead of Arg-429. (B) Subcellular localization of GFP proteins fused to either wild-type C5 sequence (a) or mutant C5-R429A sequence (b). Cells were stained as in Fig. 1B.

terminal 24 residues of Sam68 are necessary and sufficient for nuclear localization of Sam68.

Close inspection of the COOH-terminal 24 amino acids of Sam68 revealed no significant homology to any other known nuclear localization sequence. There are several basic residues in this segment, but these are rather sparsely spread throughout the entire region and do not resemble the characteristic clustering of basic amino acids seen in other nuclear localization sequences. However, direct amino acid sequence comparison of this segment to other proteins having KH domains revealed some significant homology. A short stretch of amino acids containing a minimum of two-consecutive proline residues followed by an arginine residue (PPXXR; P, Pro; R, Arg) is also conserved in the *C. elegans* germ line-specific tumor suppressor GLD-1 and several heterogenous nuclear ribonucleoproteins including hnRNP-C, hnRNP-U, hnRNP-K and hnRNP-M4. In addition, when the protein sequence data base was searched using the BLASTP program, viral transcription factors encoded by the VP16 gene of Herpes virus were found to contain related sequences. In these proteins the first proline and last arginine of the PPXXR motif tend to be preceded by alanine or proline.Fig. 2

To demonstrate the importance of the PPXXR motif, we examined the nuclear targeting activity of a mutated version of GFP-C5 (GFP-C5-R429A) in which Arg-429, the last residue of the PPXXR motif, was replaced with Ala (Fig. 3A). The accumulation of this mutant fusion protein in the nucleus was significantly diminished (Fig. 3B), suggesting the prime importance of this conserved residue in nuclear translocation.

## 4. Discussion

In this study, we have delineated a sequence necessary for localizing Sam68 to the nucleus. Removal of this sequence from the protein resulted in the localization of the mutant protein in the cytoplasm. In addition, fusion of this sequence to a heterologous protein resulted in the accumulation of the chimeric protein in the nucleus. From the analysis of various deletion mutants, the region required for this activity was mapped to amino acids 420–443 of Sam68. Although this region bears no apparent homology to any previously-reported nuclear localization signal, a particular motif, the PPXXR motif, consisting of a minimum of two consecutive proline residues followed by an invariable arginine residue was found to be conserved in a number of RNA binding proteins. A point mutation changing the invariable arginine in this motif to alanine abolished its nuclear accumulating activity. These results suggest that the COOH-terminal 24 amino acids of Sam68 function as a nuclear localization signal.

The most intensively-studied nuclear localization signal to date is the one found in the SV40 large T antigen in which a minimal sequence of seven amino acids PKKKRKV was found to be sufficient for targeting the pyruvate kinase fusion protein to the nucleus [16]. Another well-studied nuclear localization signal is the bipartite nuclear localization sequence, the prototype of which is the one found in nucleoplasmin. In the bipartite nuclear localization sequence, two basic amino acids are followed by a spacer region of any ten amino acids and then a cluster of five basic amino acids in which the three out of five must be basic residues [17]. In addition, a unique nuclear localization signal, which confers a shuttling activity to fused proteins was recently identified within hnRNP-A1, one of a heterogenous group of ribonucleoproteins that constantly shuttle between the nucleus and the cytoplasm. The sequence responsible for the shuttling activity of hnRNP-A1 has been delineated to a 38 amino acid segment [18]. The novel nuclear localization signal of Sam68 reported in this study, however, is not similar to any of these previously reported sequences, and thus defines a novel class of nuclear localization sequence.

Since protein-protein interactions have been shown to play an important function in the process of nuclear import [20,21], the conserved arginine residue in the PPXXR motif may be directly involved in the interaction with components or regulators of the nuclear import machinery. Interestingly, it has recently been demonstrated that Sam68 interacts with the 3D RNA-dependent RNA polymerase of poliovirus and is relocated from the nucleus to the cytoplasm in cells infected with poliovirus [22]. Although the precise mechanism of this relocation is not known, binding of viral proteins to Sam68 could conceivably disrupt the interaction between Sam68 and the cellular protein machinery that normally keeps Sam68 in the nucleus. Alternatively, there is a possibility that the motif may be functioning as an RNA binding domain and that Sam68 may be imported into the nucleus with RNA molecules such as snRNAs which re-enter the nucleus [23].

Whether this conserved motif is sufficient for Sam68 nuclear localization and whether this motif is also functional in proteins other than Sam68 are currently under study in our laboratory. Interestingly, the recently-characterized nuclear localization and retention signal for hnRNP-C1, which was delineated to a minimal 58 amino acids residues, contains this consensus sequence [19]. However, in hnRNP-C1, this motif may in fact function as a part of the nuclear retention signal because a region immediately downstream of the consensus sequence was shown by deletion analysis to be more important for the nuclear localization of the protein.

The presence of this conserved sequence in various RNA binding proteins suggests that this motif is a nuclear localization signal commonly utilized by a subset of RNA binding proteins of common cellular function. hnRNP proteins in the nucleus bind to nascent pre-mRNA as it is transcribed and are believed to be involved in the maturation of mRNA. In some cases, such as hnRNP-A1 and hnRNP-K, which constantly

shuttle between the nucleus and the cytoplasm, they are most likely involved in mRNA export from the nucleus. Sam68 may also be involved in each of these steps. Identification of the nuclear localization signal of Sam68 would facilitate our understanding of the mechanism that regulates the subcellular localization of Sam68 and may provide clues as to how the nuclear functions of Sam68 are coupled to the Src signal transduction pathway.

## References


[1] S.J. Taylor, D. Shalloway, Nature 368 (1994) 867–871.
[2] S. Fumagalli, N. Totty, J.J. Hsuan, S.A. Courtneidge, Nature 368 (1994) 871–874.
[3] S.J. Taylor, M. Anafi, T. Pawson, D. Shalloway, J. Biol. Chem. 270 (1995) 10120–10124.
[4] S. Richard, D. Yu, K.J. Blumer, D. Hausladen, M.W. Olszowy, P.A. Connelly, A.S. Shaw, Mol. Cell. Biol. 15 (1995) 186–197.
[5] G.M. Twamley-Stein, R. Pepperkok, W. Ansorge, S.A. Courtneidge, Proc. Natl. Acad. Sci. USA. 90 (1993) 7696–7700.
[6] S. Roch, M. Koegl, M. Vittoria Barone, M.F. Roussel, S.A. Courtneidge, Mol. Cell. Biol. 15 (1995) 1102–1109.
[7] S. Roche, S. Fumagalli, S.A. Courtneidge, Science 269 (1995) 1567–1569.
[8] P. Lock, F. Fumagalli, P. Polakis, F. McCormick, S.A. Courtneidge, Cell 84 (1996) 23–24.
[9] G. Wong, O. Muller, R. Clark, L. Conroy, M.F. Moran, P. Polakis, F. McCormick, Cell 69 (1992) 551–558.
[10] H. Siomi, M.J. Matunis, W.M. Michael, G. Dreyfuss, Nucleic Acids Res. 21 (1993) 1193–1198.
[11] M. Cruz-Alvarez, A. Pellicer, J. Biol. Chem. 262 (1987) 13377–13380.
[12] H. Siomi, M.C. Siomi, R.L. Nussbaum, G. Dreyfuss, Cell 74 (1993) 291–298.
[13] A.R. Jones, T. Schedl, Genes and Dev. 9 (1995) 1491–1504.
[14] M. Kiledjian, G. Dreyfuss, EMBO J. 11 (1992) 2655–2664.
[15] L. Wang, S. Richard, A.S. Shaw, J. Biol. Chem. 270 (1995) 2010–2013.
[16] D. Kalderon, B.L. Roberts, W.D. Richardson, A.E. Smith, Cell 39 (1984) 499–509.
[17] C. Dingwall, et al. J. Cell. Biol. 107 (1988) 841–849.
[18] H. Siomi, G. Dreyfuss, J. Cell. Biol. 129 (1995) 551–560.
[19] S. Nakielny, G. Dreyfuss, J. Cell. Biol. 134 (1996) 1365–1373.
[20] D. Görlich, I. Mattaj, Science 271 (1996) 1513–1518.
[21] V.W. Pollard, W.M. Michael, S. Nakielny, M. Siomi, F. Wang, G. Dreyfuss, Cell 86 (1996) 985–994.
[22] McBride, A.E., Schlegel, A. and Kirkegaard, K. (1996) Proc. Natl. Acad. Sci. USA 93, 2296–2301.
[23] N. Michaud, D. Goldfarb, J. Cell. Biol. 116 (4) (1992) 851–861.